基于深度时空卷积神经网络的人群异常行为检测和定位 *

胡学敏，陈 钦，杨 丽†，余 进，童秀迟

(湖北大学计算机与信息工程学院，武汉 430062)

摘 要：针对公共场合人群异常行为检测准确率不高和训练样本缺乏的问题，提出一种基于深度时空卷积神经网络的人群异常行为检测和定位的方法。首先针对监控视频中人群行为的特点，综合利用静态图像的空间特征和前后帧的时间特征，将二维卷积扩展到三维空间，设计面向人群异常行为检测和定位的深度时空卷积神经网络；为了定位人群异常行为，将视频分成若干子区域，获取视频的子区域时空数据样本，然后将数据样本输入设计的深度时空卷积神经网络进行训练和分类，实现人群异常行为的检测与定位。同时，为了解决深度时空卷积神经网络训练时样本数量不足的问题，设计一种迁移学习的方法，利用样本数量多的数据集预训练网络，然后在待测试的数据集中进行微调和优化网络模型。实验结果表明，该方法在 UCSD 和 Subway 公开数据集上的检测准确率分别达到了 99%和 93%以上。

关键词：人群异常行为检测；深度时空卷积神经网络；迁移学习；数据扩充

中图分类号：TP391.41 doi: 10.19734/j.issn.1001-3695.2018.09.0671

Abnormal crowd behavior detection and localization based on deep spatial-temporal convolutional neural networks

Hu Xuemin, Chen Qin, Yang Li†, Yu Jin, Tong Xiuchi

(*School of Computer Science & Information Engineering, Hubei University, Wuhan* 430062, *China*)

Abstract: To handle the issues of low accuracy and lacking training samples in abnormal crowd behavior detection in public places, this paper proposes a method based on deep spatial-temporal convolutional neural networks in this paper. In view of the characteristics of crowd behavior in monitoring videos, a deep spatial-temporal convolution neural network for detecting abnormal crowd behavior is first designed by extending 2D convolution to the 3D space according to spatial features of static images and temporal features between the frames before and after the current frame. To locating abnormal crowd, this paper divides video frames into a number of subregions that obtain spatial-temporal samples. Then, the samples are input into the designed deep spatial-temporal convolutional neural network for training and classification, whose results are used to detect and locate abnormal crowd. In the meanwhile, this paper utilizes a transfer learning method to deal with the issue of lacking training samples when training the deep spatial-temporal convolutional neural network, where datasets with more training samples are used to pre-train the network which is fine-tuned and optimized on testing datasets with fewer samples. Experimental results show that the detection accuracies on UCSD and Subway open datasets are greater than 99% and 93%, respectively.

Key words: crowd abnormal behavior detection; deep spatial-temporal convolutional neural network; transfer learning; data augmentation

0 引言

近年来，随着城市人群安全问题日益突出，视频监控显得尤为重要。传统视频监控通过工作人员观察监控画面获知异常情况，这种方法不仅主观性强，而且浪费人力、效率低下。因此，关于人群异常行为检测与定位的智能视频监控系统具有重要研究意义和商业价值。

目前，国内外研究人员已经在人群异常检测方面做了不少研究工作，并取得了一定成果。相关方法主要分为两大类。第一类是基于局部目标检测的方法，该方法通常利用动态模型对人群行为进行建模。Chaudhry 等人[1]利用面向对象的光学直方图并结合分类器来识别人群行为；Colque 等人[2,3]利用基于方向、速度和熵的直方图描述人群的异常；Li 等人[4]则提出一种动态混合纹理模型来实现人群异常的检测。这类方法能够有效检测并定位异常人群行为，但是模型构建复杂并且检测率不高。第二类则是基于全局统计的方法，从整体提取某些特征，如角点、梯度、光流等，然后通过特征分类的方法来实现人群异常行为检测。王乔等人[5]提出一种基于整体能量模型表示的方法来较辨识正常行为中的异常行为；任晓芳等人[6]对输入的视频使用梯度方向直方图特征和光流直方图特征识别人体动作,最后结合基于能量的最小二乘双分界面支持向量机完成人体动作的识别；姬丽娜等人[7]提出一种基于混合高斯模型和尺度不变特征变换特征的人群数量统计分析方法。这类算法的缺点是受环境中光线影响比较大，

收稿日期：2018-09-08；**修回日期：**2018-10-25 **基金项目：**国家自然科学基金青年基金资助项目（61806076）；湖北省自然科学基金青年资助项目（2018CFB158）；湖北省大学生创新创业训练计划基金资助项目（201710512049）；湖北省人文社科重点研究基地开放课题（2015JX04）

作者简介：胡学敏（1985-），男，湖南岳阳人，副教授，博士，主要研究方向为计算机视觉和智能系统；陈钦（1995-），男，湖北黄冈人，本科生，主要研究方向为计算机视觉；杨丽（1985-），女（通信作者），山西长治人，讲师，博士，主要研究方向为智能计算（45693296@qq.com）；余进（1998-），男，湖北天门人，本科生，主要研究方向为机器学习；童秀迟（1996-），女，湖北随州人，硕士研究生，主要研究方向为智能系统.

准确率不高。

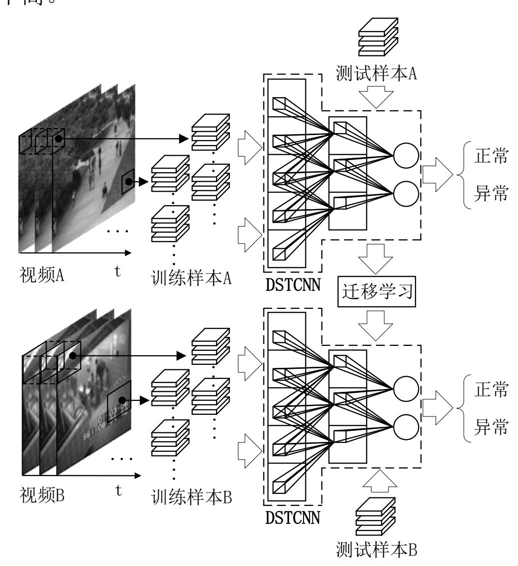


图 1　基于 DSTCNN 的人群异常行为识别与定位流程图

Fig. 1　Flow chart of abnormal crowd behavior detection and localization based on DSTCNN

近年来，卷积神经网络(convolution neural network, CNN)凭借其优良的特征提取能力在计算机视觉领域得到广泛关注。传统卷积神经网络能够对二维图像进行有效的特征提取，从而实现目标检测与分类。Chen 等人[8]借助卷积神经网络实现在夜景中识别汽车转向灯；文献[9]则利用卷积神经网络对人的年龄和性别进行判断。传统卷积神经网络仅能够在二维图像上提取特征，但是无法应用于三维的视频数据。文献[10]提出了基于传统 CNN 的人群异常行为检测，但是只使用了运动方向、速度和加速度三者的运动显著图，丢失了大量信息，导致能够检测的异常行为有限。针对这种问题，有学者提出能够应用于视频中时空特征提取的时空卷积神经网络，并应用于与人体行为的识别[11]。此外，训练一个深度网络需要大规模、多样化的训练样本，而实际的人群异常行为检测中往往难以获取足够的样本，从而导致检测效果不理想。

针对现有的人群异常检测方法检测率不高、传统 CNN 无法提取时间相关特征、以及训练样本缺乏的问题，本文提出一种基于深度时空卷积神经网络(Deep Spatial-temporal Convolution Neural Network, DSTCNN)的人群异常行为检测与定位的方法，如图 1 所示。首先，基于传统 CNN，结合时间特征，设计 DSTCNN 的结构；然后基于 DSTCNN，设计人群异常行为检测和定位的方法；针对数据不足问题，提出基于迁移学习的 DSTCNN 训练方法。在数据量较多的数据集上训练得到检测率较高的 DSTCNN 模型，将此模型通过迁移学习(transfer learning, TF)的方法迁移到其他数据集对应模型上并训练。实验结果表明本文的方法与现有方法相比，具有更高的检测率。

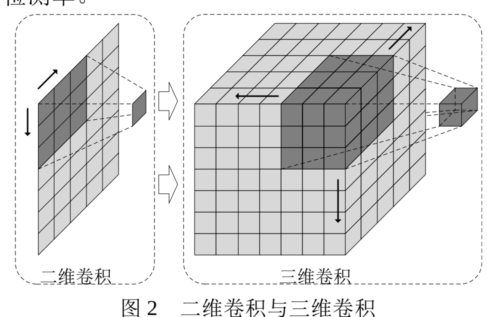


图 2　二维卷积与三维卷积

Fig. 2　2D convolution and 3D convolution

1　时空卷积神经网络

传统卷积神经网络使用二维卷积核对图像进行特征提取，卷积神经网络中每个卷积层都包含多个不同的卷积核，每个卷积核提取不同的特征。通常一个完整的卷积神经网络包含多层卷积层，低层卷积层提取低级特征，如边缘、线条、角落；高层卷积层提取高级特征，如目标对象各部分的轮廓。卷积层提取的特征最后通过全连接层组合成完整的目标对象。传统卷积神经网络可以有效地对图像进行特征提取和目标分类，但是仅限于静态图像，无法对视频中连续视频帧之间的时间特征进行提取。时空卷积神经网络和传统卷积神经网络的本质区别在于网络所使用卷积核的不同，如图 2 所示。传统卷积神经网络中所使用的二维卷积核提取特征时，只在图像上进行行和列的卷积。而连续视频帧中，除了每一帧的图像特征，连续帧之间还存在时间关联性。基于视频分析的人群异常行为检测中，人群的行为特征在视频中表现为空间和时间的关联性。因此时空卷积神经网络采用三维卷积核，其卷积计算的内容除了包含每一帧图像的行和列像素点以外，还包含前后帧对应位置的像素点，即时空卷积神经网络除了能够提取单个视频帧的图像特征，还能提取连续帧之间的时间特征。时空卷积神经网络的每一层从输入到输出的计算方式如式(1)所示。

$$y=\sigma\left(\sum_{k}\sum_{j}\sum_{i}(\vec{x}_{ijk}*\vec{w}+b)\right) \tag{1}$$

其中：y 表示某一层的输出；σ 表示激活函数；i、j、k 表示样本上对应位置的坐标；$\vec{x}_{ijk}$ 表示每一层输入上对应于(i , j , k)处与对应卷积核尺寸大小相等的局部区域，如图 2 右边立方体中阴影部分所示；$\vec{w}$ 表示卷积核的权重矩阵；b 表示对应卷积核的偏置值。

2　人群异常行为检测与定位算法

为实现人群异常的定位，本文首先将完整的视频画面分成若干子区域并编号，对每个子区域使用 DSTCNN 进行人群异常行为检测，若某个子区域检测为异常，则可根据编号找到对应的视频区域，实现人群异常行为的检测与定位。

2.1　深度时空卷积神经网络结构设计

本文设计的面向人群异常行为检测和定位的深度时空卷积神经网络结构，如图 3 所示，其中包含 8 个卷积层、5 个池化层和 2 个全连接层以及 1 个输出层。为提取人群行为信息，以连续若干帧一定大小视频子区域作为输入。由于本文设计深度时空卷积神经网络主要目的在于检测异常，因此输出为两类，正常与异常。

实验证明 3×3×3 的卷积核尺寸对于视频处理是一种合适的尺寸[12]，因此本文中的深度时空卷积神经网络的卷积核尺寸统一固定为 3×3×3；卷积步长统一设定为 1×1×1。除第一层池化核大小设定为 1×2×2，步长设定为 1×2×2 以外，其他层的池化核大小统一设定为 2×2×2，步长设定为 2×2×2。池化方式统一采用最大池化，并将池化层间隔一层或两层置于卷积层中间，对卷积层信息降采样，用以在保留重要信息和减少相关度相对较低信息、降低计算复杂度的同时，提升卷积神经网络的泛化能力。而第一层池化层 1×2×2 池化核与池化步长的设计可以避免视频中时序信息被过早地降采样。第 1~4 层卷积层用于提取低级特征，第 5~8 层卷积层用于提取高级特征，低级特征较为通用，种类较少，高级特征更为具体，种类较多，因此第 1 层卷积层设置 64 个卷

积核，第 2 层卷积层设置 128 个卷积核，第 3~5 层卷积层设置 256 个卷积核，第 6~8 层卷积层设置 512 个卷积核。通过卷积与池化得到的特征与两层全连接层连接组成具体目标对象，最后通过输出层得到卷积神经网络对输入视频的分类概率。

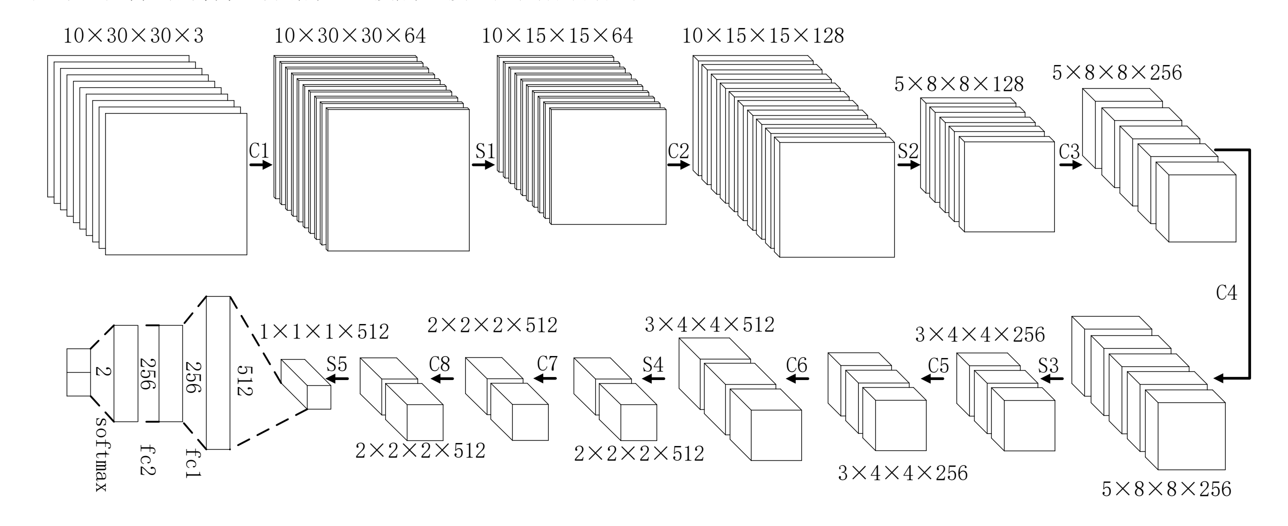


图 3　面向人群异常行为检测和定位的深度时空卷积神经网络结构

Fig. 3　Structure of deep spatial-temporal convolutional neural network for crowd abnormal behavior detection and localization

在每一层卷积层以及全连接层之后，本文使用 ReLu(Rectified linear unit,修正线性单元)激活函数，提升神经网络模型的线性表达能力。为输出视频的分类概率，输出层采用 Softmax 函数。最后结合交叉熵[13]函数与 L2 正则化构造损失函数，并利用 Adam 优化算法更新优化神经网络模型。交叉熵函数如式(2)所示。

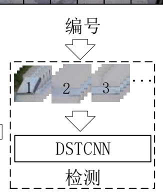
$$C=-\frac{1}{n}\sum_{x}[y\ln a+(1-y)\ln(1-a)] \tag{2}$$

其中：C 表示损失值；n 表示样本总量；x 表示样本；y 表示期望的输出(即标签)；a 表示实际的输出，具体表达式如式(3)所示。

$$a=\sigma\left[\sum_{k}\sum_{j}\sum_{i}(\bar{x}_{ijk}*\bar{w}+b)\right] \tag{3}$$

划分局部子区域

编号





定位

图 4　异常检测与定位

Fig. 4　Anomaly detection and location

2.2　异常行为的检测与定位

为实现人群异常行为的定位，本文将完整视频画面划分为若干个大小相同的局部子区域，相邻子区域互不重叠。每个子区域设置一个唯一编号，每个编号对应不同的子区域，如图 4 所示。为实现对每个子区域进行人群异常行为检测，本文将每个子区域的连续视频帧作为输入，训练得 DSTCNN 模型；在检测人群异常行为时，同样将完整视频划分为局部子区域，再将每个局部子区域输入已训练好的 DSTCNN 模型。若某个子区域被识别为异常，则根据编号确定其在视频画面中的位置，实现人群异常行为的定位。

3　深度时空卷积神经网络的训练方法

深度神经网络的训练和优化需要大规模、多样化的数据样本，而对于人群异常行为分析，异常情况通常以小概率发生，因此难以获取足够的训练样本，特别是异常行为的样本。现有的公开数据集中，同样存在异常行为样本数量较少、正常和异常样本数量相差较大的问题，导致难以训练实用的深度神经网络。针对这种情况，本文设计一种基于数据扩充和迁移学习的深度神经网络训练方法，在训练 DSTCNN 之前首先将训练样本数据进行数据扩充，使用扩充后的训练样本训练得到优化后的 DSTCNN 模型，然后使用迁移学习的方法，将此模型迁移到其他 DSTCNN 模型上，实现少量样本的训练和优化。

3.1　数据扩充

图像的亮度、对比度、噪声等属于图像的二维特征，且这些特征对于人群的行为没有影响，因此本文针对训练样本数据不足及正常样本和异常样本数量相差较大问题，通过增加对比度、降低对比度、增加亮度、降低亮度、添加椒盐噪声、进行高斯模糊等 6 种方式，对训练样本进行数据扩充(data augmentation, DA)使其数量增加为原来的 7 倍，如图 5 所示。同时，通过保留数量较少的异常样本并随机去除数量相对较多的正常样本，使正常样本与异常样本比例为 2：1。利用以上方式，可以改变视频中每一帧的图像信息，增加样本多样性，同时保留人群原有的行为特征，并减小正常样本与异常样本数量差异。从图 5 中可以看出，本文的数据扩充方法能够在一定程度上有效增加训练数据的数量。

3.2　基于迁移学习的 DSTCNN 训练方法

对于一些人群异常行为检测的场景或者公开数据集，其中异常样本数据过于稀少，即使在经过数据扩充后样本数据仍然不足。另外，如果只是通过数据扩充来实现正负样本的平衡，容易引起过拟合的问题。因此，本文提出一种基于迁移学习的训练方法，实现 DSTCNN 的训练和优化。

当源域数据与目标域数据存在部分共享模型参数时，即两个数据集之间存在部分相似的基础特征，而该部分特征可

利用相同卷积神经网络进行提取的时候，可基于模型进行迁移学习[14]。Long 等人[15]提出一种 DAN 结构，通过迁移并固定卷积神经网络低层卷积层网络参数，在特定数据集上微调高层卷积层及全连接层的方法，在 Office-31、Office-10 + Caltech-10 数据集上取得较高识别率。由于不同的人群行为数据集中，人群行为也存在部分相似的特征，因此本文在数据量相对较多的 UCSD 数据集上训练得到 DSTCNN 模型，再将此模型通过迁移学习迁移到 Subway 数据集上的 DSTCNN 模型，如图 6 所示。

在进行迁移学习过程中，本文将 UCSD 数据集作为源域 D_s，将 Subway 数据集作为目标域 D_t，其中：

$$D_s=\{x_i,y_i\}_{i=1}^{n},\ D_t=\{x_j,y_j\}_{j=1}^{n},\ X_s=X_t,Y_s=Y_t$$

x_i、x_j 分别表示第 i、j 个样本，y_i、y_j 表示对应的标签。X_s、X_t 分别表示源域及目标域的特征空间，Y_s、Y_t 分别表示源域及目标域的标签空间。

首先利用 D_s 学习得到分类器: f_1: $x_s \rightarrow y_s$ 来预测 D_s 的标签 y_s，其中 x_s 表示源域样本，f_1 表示 D_s 对应的目标函数，且 $x_s \in D_s$; 为得到: f_2: $x_t \rightarrow y_t$,即用来预测 D_t 对应标签 y_t 的分类器，其中 f_2 表示 D_t 对应的目标函数，x_t 表示目标域样本，且 $x_t \in D_t$，本文将 f_1: $x_s \rightarrow y_s$ 在一定条件下利用 D_t 重新学习，完成如下过程：

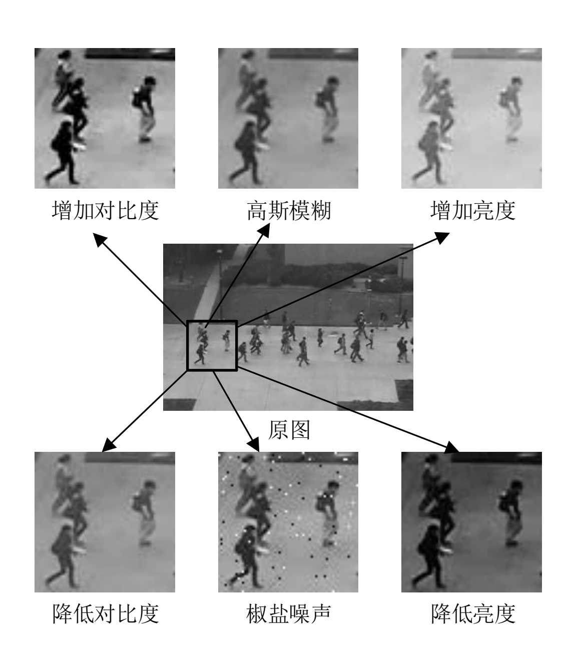


图 5 数据扩充样例

Fig. 5 Data augmentation samples

$$f_2: x_t \rightarrow y_t = \text{retrain}(f_1: x_s \rightarrow y_s)\ s.t.\ \text{Net}_1[1\sim l] = \text{Net}_2[1\sim l] = K,\ (x, y) \in D_t$$

其中：retrain 表示重新训练过程，$\text{Net}_1[1\sim l]$、$\text{Net}_2[1\sim l]$分别表示 D_s、D_t 对应的神经网络的第 1 层到第 l 层参数，K 表示常数矩阵，重新训练过程中 K 不变。为确定 l 值，本文分别对 l 取不同值并在 UCSD 数据集上进行对比实验，最终确定最优值 l=4。

训练网络时，在数据扩充的基础上，本文将两个数据集划分局部子区域，首先在 UCSD 数据集上训练 DSTCNN，取连续 N(本文中 N=10)帧 UCSD 数据集局部子区域视频作为输入，初始化 DSTCNN 模型参数后进行训练，再利用 UCSD 数据集测试集对 DSTCNN 模型进行评测，得到优化后的 DSTCNN 模型；由于在 UCSD 数据集与 Subway 数据集中，DSTCNN 的前 4 层提取的特征是两个数据集所共有的特征，因此可利用相同结构神经网络进行特征提取，即将在 UCSD 数据集上训练得到的 DSTCNN 模型前 4 层卷积层直接迁移到 Subway 数据集的 DSTCNN 上，固定该 4 层网络参数，取连续 N 帧 Subway 数据集的局部子区域视频作为输入，更新除前 4 层卷积层之外的参数，得到适合 Subway 数据集人群异常行为检测和定位的 DSTCNN 模型。

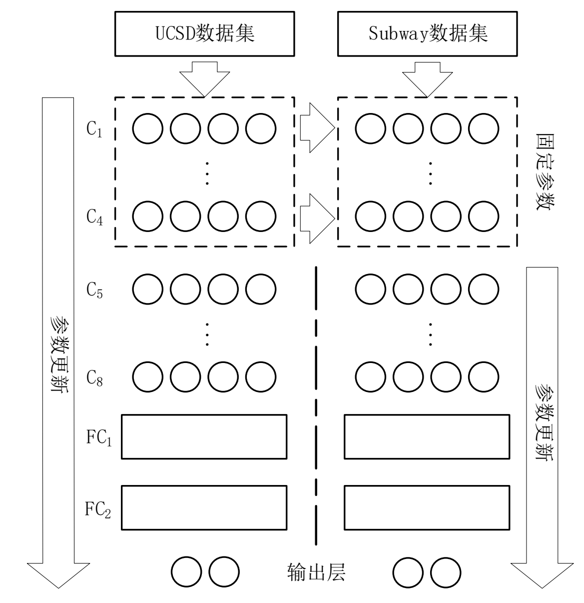


图 6 UCSD 数据集到 Subway 数据集的 DSTCNN 模型迁移学习

（C_1~C_8:第 1~8 个卷积层,FC_1~ FC_2:2 个全连接层）

Fig. 6 Transfer Learning of DSTCNN model from UCSD dataset to Subway dataset

4 实验结果与分析

本文进行实验的硬件环境：CPU 为 Core i7-7700K (Quad-core 4.2 GHz) 、显卡为 NVIDA GTX 1080ti、内存为 32 GB。软件环境：计算机操作系统为 Windows 10 pro、DSTCNN 训练测试平台为 TensorFlow1.2、Python3.5。

本文利用 UCSD 以及 Subway 这两个开源的数据集。其中 UCSD 数据集由美国加利福尼亚州统计学习视觉计算实验室提供[16]。该数据集包含场景一和场景二两个人群行走视频，其视频像素分辨率分别为 238×158、360×240。视频中正常行为是正常行走的人群，而异常的行为包括骑自行车、驾驶汽车、踩滑板、坐轮椅的老人等行为；Subway 数据集源自于 Adam 等人，包含地铁入口以及出口的两个监控视频，像素分辨率均为 512×384[17]。对于入口视频，正常行为是正常刷卡进入入口，异常行为表现为从入口出来、不刷卡强行翻越进入入口等；对于出口视频，正常行为则是正常从出口走出来，异常行为则是强行从出口翻越进入。

本文将 UCSD 及 Subway 数据集分为训练数据以及测试数据两部分，为保证训练时异常样本的数量，本文选取异常行为相对较多的视频作为训练数据，其余的作为测试数据。为了在尺寸上能够将人群中的个体、人群与障碍物分离，并且最大限度包含人体行为信息，UCSD 数据集划分的子区域大小为 30×30 像素，Subway 数据集划分的子区域大小为 60×60 像素，同一子区域取连续 10 帧作为一个样本。当视频中出现异常情况时，所有包含人体的子区域标注为异常样本，其余情况子区域均标注为正常样本。对训练数据按照 3.1 节中的方法进行预处理，得到训练数据，数据扩充前后样本数量、正常与异常样本比例如表 1 所示。可以看出，Subway

数据集中包含的人群异常行为的样本，远远无法达到训练一个深度神经网络的要求。

为证明扩充数据以及迁移学习的有效性，本文在 UCSD 以及 Subway 两个数据集的每一个场景上都进行了 4 组实验：不进行数据扩充直接训练 DSTCNN、进行数据扩充后训练 DSTCNN、不进行数据扩充并结合迁移学习训练 DSTCNN、进行数据扩充后结合迁移学习训练 DSTCNN。其中，对 UCSD 数据集，其场景一和场景二样本相似度较大，因此训练时合并两个场景一起训练，但测试时分场景一和场景二两个场景单独测试。UCSD 数据集实验结果如图 7(a)-(d)所示，图中黑色方框区域即为异常区域；对 Subway 数据集，其出口和入口两个场景样本差异较大，因此分开训练及测试。Subway 数据集的实验结果如图 8(a)-(d)所示。为定量描述实验结果，本文基于每个测试数据集绘制受试者工作特性曲线(Receiver Operating Curve, ROC)，并计算出 ROC 曲线下的面积(Area Under the Curve, AUC)，UCSD 数据集的场景一、场景二以及 Subway 数据集的入口、出口场景 ROC 曲线分别如图 7(e)(f)、图 8(e)(f)，其 AUC 的值如表 2 所示。另外，为验证本文方法的有效性，本文将实验结果与几种经典算法 HOOF[1]、HOFM[2]、HOFME[3]、MDT-temporal[4]、MDT-spatial[4]进行对比，结果如表 3 所示。根据实验结果，可以得出以下结论：

a）本文设计的 DSTCNN 模型能够有效检测和定位多种人群异常行为。DSTCNN 通过提取二维图像特征和视频序列特征，表达复杂的人群行为，因此本文方法能够检测和定位多种不同的人群异常行为。如图 7(a)中检测出的汽车、图 8(a)中检测出的翻越地铁入口的人群。

b）本文提出的数据扩充和基于迁移学习的训练方法，能够有效提高人群异常行为检测的准确率。本文通过数据扩充增加了训练样本的数量以及多样性，同时结合迁移学习优化 DSTCNN 模型，使其即使在训练样本数量很少的情况下，也能够准确提取人群行为特征并分类，从而提高了人群异常行为检测的准确率。例如表 3 中 UCSD 数据集的场景一通过数据扩充以及迁移学习的方法将 AUC 值提升了 0.1663；表 3 中 Subway 数据集的入口场景 AUC 值通过相同方法提升了 0.0539。

c）本文基于 DSTCNN，结合数据扩充与迁移学习的方法，比现有经典方法在公开数据集中的检测率更高。本文分别通过数据扩充与迁移学习在数据层面及模型层面对 DSTCNN 进行优化，最终在 UCSD 数据集与 Subway 数据集上得到的测试结果 AUC 均高于对比的几种经典方法，其中 UCSD 数据集场景二的 AUC 更是达到了 0.9994，相比对比方法中最好的 AUC 为 0.899 高出 0.1004。此外，在正常和异常人群样本数量悬殊的情况下，只进行数据扩充训练神经网络模型比只进行迁移学习效果好。迁移学习能够解决因数据量较少导致的深度神经网络性能不高的问题，但是无法有效处理数据正负样本数量悬殊的问题。本文对数据进行扩充后对正常和异常人群的样本比例进行了调整，因此数据扩充后训练得到的模型性能优于仅使用迁移学习训练得到的深度神经网络模型。

表 1　不同数据集的样本数量

Tab.1 Sample numbers of different datasets

数据集	场景	有无扩充	训练		测试	
			正常样本	异常样本	正常样本	异常样本
UCSD	场景一	无	1332	666	1388	12
		有	9324	4662	—	—
	场景二	无	1332	666	3398	58
		有	9324	4662	—	—
Subway	入口	无	164	82	2796	36
		有	1148	574	—	—
	出口	无	108	54	2396	6
		有	756	378	—	—



(a)场景一(汽车)

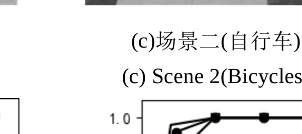
(a) Scene 1(cars)





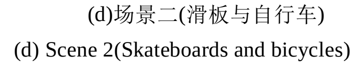
(b)场景一(轮椅)

(b) Scene 1(Wheelchairs)



(c)场景二(自行车)

(c) Scene 2(Bicycles)

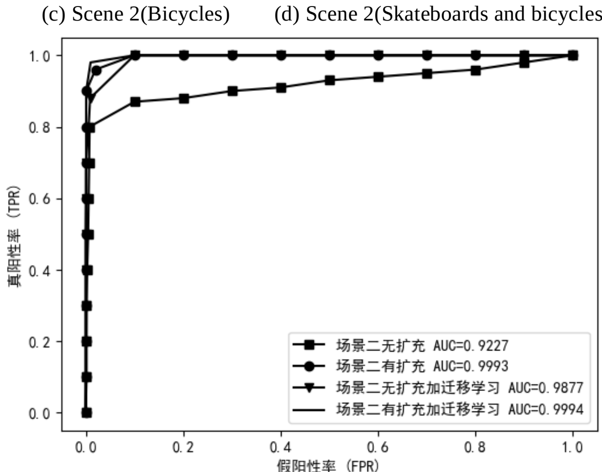


(d)场景二(滑板与自行车)

(d) Scene 2(Skateboards and bicycles)

(e) UCSD 场景一 ROC 曲线及 AUC

(e) ROCs and AUCs of Scene 1 in UCSD dataset



(f) UCSD 场景二 ROC 曲线及 AUC

(f) ROCs and AUCs of Scene 2 in UCSD dataset

图 7　UCSD 数据集测试结果

Fig. 7　Experimental results in UCSD dataset





(a)入口场景(一)

(a) Entrance (1)



(b)入口场景(二)

(b) Entrance (2)

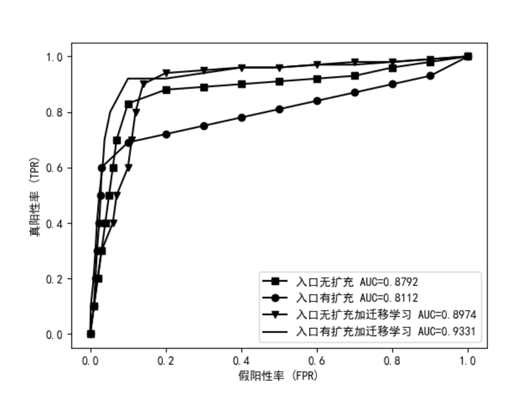


(c)出口场景(一)

(c) Exit (1)

(d)出口场景（二）

(d) Exit (2)



(e) Subway 入口场景 ROC 曲线及 AUC

(e) ROCs and AUCs of Entrances in Subway dataset

(f) Subway 出口场景 ROC 曲线及 AUC

(f) ROCs and AUCs of Exits in Subway dataset

图 8　Subway 数据集测试结果

Fig. 8　Experimental results in Subway dataset

d）在 8 层卷积层的深度时空卷积神经网络中，固定前 4 层时空卷积层进行迁移学习的能取得最优检测效果。表 4 为将在 UCSD 数据集上训练的网络，迁移到 Subway 入口数据集上训练，固定前 1~l 层测试的 AUC 值。由于第 4 层之前的时空卷积层主要提取的为行人的边缘、形状等人群通用特征，后 4 层的时空卷积层主要提取为行人的行为特征，因此固定前 4 层的检测效果较好。随着固定层数的减少，由于 Subway 上的数据量较少，少量的数据难以训练所有网络层的权重，故 AUC 值会减小；随着固定层数的增加，由于两个数据集的人群异常行为的专用特征不一样，因此只微调个别高层网络的权重，无法拟合新的数据集的人群行为特征数据，因此 AUC 值也会减小。

表 2　有无数据扩充和迁移学习的对比测试

Tab.2 Comparative results between experiments with DA or TF and experiments without DA or TF

方法	UCSD 数据集		Subway 数据集	
	场景一	场景二	入口	出口
DSTCNN	0.8311	0.9227	0.8792	0.5000
DSTCNN+DA	0.9960	0.9993	0.8112	0.8984
DSTCNN+TF	0.9919	0.9877	0.8974	0.9820
DSTCNN+DA+TF	0.9974	0.9994	0.9331	0.9829

表 3　与其他经典方法的对比测试

Tab.3 Comparative results with other classical methods

方法	UCSD 数据集		Subway 数据集	
	场景一	场景二	入口	出口
HOOF	0.6900	0.8200	0.7740	0.8000
HOFM	0.7150	0.8990	0.8150	0.8450
HOFME	0.8490	0.8160	0.8160	0.8490
MDT-temporal	0.8250	0.7650	0.8890	0.8750
MDT-spatial	0.6000	0.7500	0.6820	0.6700
DSTCNN+DA+TF	0.9974	0.9994	0.9331	0.9829

表 4　不同迁移层数在 Subway 入口数据集上测试 AUC 值

Tab.4 AUC values of testing results for different transfer layers on the Subway dataset of the entrance

迁移层数 1~l 的 l 取值	2	3	4	5	6
AUC	0.8227	0.8883	0.9331	0.9068	0.8734

5　结束语

本文提出一种利用深度时空卷积神经网络，并结合迁移学习实现人群异常检测与定位的方法。该方法中，首先根据应用场景设计 DSTCNN 结构，该结构主要包含用于特征提取的卷积层与特征分类的全连接层及输出层，然后设计基于数据扩充和迁移学习的训练方法，实现 DSTCNN 的训练和优化，提高检测率。在 UCSD 数据集和 Subway 数据集上的测试结果表明，本文的方法能够有效进行人群异常行为检测与定位，其检测准确率高于几种经典方法。

同时，本文方法也存在一定局限性。由于 DSTCNN 计算量较大，本文的方法实时性难以满足实时性要求高的多路监控系统；此外，本文方法只能检测并定位异常，无法识别出具体是何种异常。因此未来的工作将致力于优化算法，提高实时性以及实现异常行为的分类识别。

参考文献：

[1] Chaudhry R, Ravichandran A, Hager G, *et al.* Histograms of oriented optical flow and Binet-Cauchy kernels on nonlinear dynamical systems for the recognition of human actions [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2009: 1932-1939.

[2] Colque R V H M, Junior C A C, Schwartz W R. Histograms of Optical Flow Orientation and Magnitude to Detect Anomalous Events in videos [C]//Proc of the 28th SIBGRAPI Conference on Graphics, Patterns and Images. Washington DC: IEEE Computer Society, 2015: 126-133.

[3] Colque R V H M, Junior C A C, Schwartz W R. Histograms of optical flow orientation and magnitude to detect anomalous events in videos [J]. IEEE Trans on Circuits & Systems for Video Technology, 2017, 27(3): 673-682.

[4] Li Weixin, Mahadevan V, Vasconcelos N. Anomaly detection and localization in crowded scenes [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2014, 36(1): 18-32.

[5] 王乔，雷航，郝宗波．基于整体能量模型的异常行为检测 [J]．计算机应用研究，2012，29(12)：4782-4785. (Wang Qiao, Lei Hang, Hao Zongbo. Abnormal behavior detection based on globe energy model [J]. Application Research of Computers, 2014, 29(12): 4782-4785.)

[6] 任晓芳，秦健勇，杨杰，等．基于能量模型的 LS-TSVM 在人体动作识别中的应用 [J]．计算机应用研究，2016，33(2)：598-601．(Ren Xiaofang, Qin Jianyong, Yang Jie, *et al*. Energy model based LS-TSVM for action recognition [J]. Application Research of Computers, 2016, 33 (2): 598-601.)

[7] 姬丽娜，陈庆奎，陈圆金，等．基于 GPU 的视频流人群实时计数 [J]．计算机应用，2017，37 (1)：145-152. (Ji Lina, Chen Qingkui, Chen Yuanjin, *et al*. Real-time crowd counting method from video stream based on GPU [J]. Journal of Computer Applications, 2017, 37(1): 145-152.)

[8] Chen Long, Hu Xuemin, Xu Tong, *et al.* Turn Signal Detection During Nighttime by CNN Detector and Perceptual Hashing Tracking [J]. IEEE Trans on Intelligent Transportation Systems, 2017, 18 (12): 3303-3314.

[9] Levi G, Hassner T. Age and gender classification using convolutional neural networks [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition Workshops. Washington DC: IEEE Computer Society, 2015: 34-42.

[10] 胡学敏，易重辉，陈钦，等．基于运动显著图的人群异常行为检测 [J]．计算机应用，2018，38(4)：1164-1169. (Hu Xuemin, Yi Chonghui, Chen Qin, *et al.* Abnormal crowd behavior detection based on motion saliency map [J]. Journal of Computer Applications, 2018, 38 (4): 1164-1169.)

[11] Ji Shuiwang, Yang Ming, Yu Kai, *et al*. 3D convolutional neural networks for human action recognition [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2013, 35 (1): 221-231.

[12] Tran D, Bourdev L, Fergus R, *et al.* Learning Spatiotemporal Features with 3D Convolutional Networks [C]//Proc of IEEE International Conference on Computer Vision. Washington DC: IEEE Computer Society, 2015: 4489-4497.

[13] Kline M, Berardi L. Revisiting squared-error and cross-entropy functions for training neural network classifiers [M]. Berlin: Springer-Verlag, 2005: 310-318.

[14] Pan Jialin, Yang Qiang. A survey on transfer learning [J]. IEEE Trans on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.

[15] Long Mingsheng, Cao Yue, Wang Jianmin, *et al.* Learning transferable features with deep adaptation networks [C]//Proc of International Conference on Machine Learning.2015: 97-105.

[16] S. V. C. Lab, “UCSD anomaly data set” [EB/OL]. (2014) [2018-10-22]. http://www. svcl. ucsd. edu/projects/anomaly/; http://www. svcl. ucsd. edu/projects/anomaly/.

[17] Adam A, Rivlin E, Shimshoni I, *et al.* Robust Real-Time Unusual Event Detection using Multiple Fixed-Location Monitors [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2008, 30 (3): 555-560.